

11 April 1960

CENTRAL INTELLIGENCE AGENCY

The cycle of organizational activity for intelligence purposes extends from the collection of selected information to its direct use in reports prepared for policy makers. Between these beginning and end activities there lie a number of functions which can be grouped under the term information processing. These functions include the identification, recording, organization, storage, recall, conversion into more useful forms, synthesis and dissemination of the intellectual content of the information collected. The ever-mounting volume of information produced and promptly wanted and the high cost of performing these manifold operations are forcing a critical review of current practices in the processing field.

To cope with its information handling problems, the Central Intelligence Agency has over the past thirteen years developed an information processing center which comprehensively indexes and stores that information which is collected and, as a service of common concern, renders daily support to analysts at work in all parts of the U.S. Government's intelligence community. This central reference service organizationally consists of a central library of books and documents, specialized libraries or "registers" concerned with biographic, graphic and industrial information, a document center to which and from which the very extensive

documentary flow comes, and a machine unit which acts as a nucleus supporting the office through controlled manipulation of data by machine methods.

Efficient and economical storage and retrieval of information is by all odds the toughest of the information processing problems, millions are being spent on it by the research libraries of universities, of industry, and of government. For us this problem is particularly vexing since our document center alone receives thousands of different intelligence documents each week in numbers of copies running into the tens of thousands. This is exclusive of newspapers, press summaries, books, maps and other such open material which is acquired by the library in an average of 200,000 pieces per month. The open literature is obtained to meet the needs of our own analysts or those of 20 other U.S. Government agencies; that which is filed centrally in our library is handled as it would be in a conventional library, using the Library of Congress classification system.

The classified documents are received from scores of different major sources, the daily volume fluctuates and lacks uniformity in format, in reproduction media, in length and quality of presentation, and in security classification. As they come in they must be read with an eye to identifying material of interest to the many different customer receipt points; those which have future retrieval value (approximately 50%) must be indexed and stored in such manner as to provide retrieval pertinent to customer needs. This material is subject to control through IBM punched cards.

These IBM card files now contain over 40 million cards. Since 1954 we have been miniaturizing the documents by microphotography and mounting

them in apertures on IBM punched cards. Access to the document itself is indirect, through codes punched into the cards to indicate subject, area, source, security classification, date and number of each document. The data on cards retrieved in response to a particular request is reproduced by photographic means in tape form and constitutes the bibliography given the customer. This system, which we instituted in 1947 (using microfilm strips rather than aperture cards prior to 1954 and facsimile rather than photo tape prior to 1959), we call the Intellofax System; it represents pioneer work in the field of information storage and retrieval.

Foreign scientific information is a part of our total volume acquisition and is important to us in the discharge of some of CIA's direct responsibilities to the National Security Council. We are also called upon to perform certain services of common concern for the intelligence community. The problem we have concerning scientific information is fourfold: 1) knowing which scientific publications we need and which are available; 2) acquiring these as promptly as possible; 3) disseminating them or the information in them in a form and in a language facilitating their use; 4) organizing the information in such manner as to permit its rapid recall when needed.

Generally speaking, it is possible to obtain copies of all significant scientific publications published abroad; exceptions are those coming under foreign security classifications and those unclassified publications whose distribution is restricted by Communist Bloc countries. Procurement of the desired publications is undertaken by the Foreign Service of the United States with community needs coordinated through an inter-agency committee of the United States Intelligence Board. This committee

includes, in addition to intelligence community representatives, members from the three national libraries (Library of Congress, National Library of Medicine and the Department of Agriculture Library), the National Science Foundation and the United States Information Agency. Through this mechanism the needs of all are known, tasks can be allocated, and performance evaluated.

We are often asked whether we make these publications available to the public. The answer is that we do. Single-copy material which we retain is given to the Library of Congress for microfilming for its collection; duplicate copies are passed on for direct use. Material which has been exploited by our research personnel and no longer needed is redistributed and thenceforth available to users of the national libraries. Of the total publications we purchase or secure abroad, 98% is made available to the public through our cooperative arrangements with the national libraries.

The Central Intelligence Agency contributes regularly material which is incorporated into the national libraries published indexes: the Monthly List of Russian Accessions, the East European Accessions Index, the Bibliography of Agriculture, and the Current List of Medical Literature. Similarly CIA maintains a continuous program of reporting selective extracts from scientific and technical literature. This information is carried in a semi-monthly publication, the Scientific Information Report, which is made available to, and reproduced by, the Office of Technical Services, Department of Commerce, who offers it to U.S. science and industry on subscription.

To meet our internal needs, we require considerable translation work. Soviet publication in the field of science and technology has

been estimated at 300 million words per year; presently about 2% of this volume is being translated. This naturally led us to the possibilities of machine translation and since 1952 CIA has been supporting machine translation research.

25X1A

The major machine translation emphasis to date has been from Russian to English in the field of chemistry and from French to English in nuclear physics. A capability, using a general purpose computer (IBM 704 and 705), now exists for production of usable text (requiring minimal stylization) at the rate of 50,000 words per hour from Russian and 30,000 words per hour from French. Costs are comparable to human translation and, of course, speed is much greater, a human translator producing about 2600 words per day. Present problems involve input to the computer to keep it working at optimum since today the input must be hand-punched, and a key punch operator can produce between 5-6,000 cards per day. Optimum computer use therefore requires large key punch staffs. The solution to this problem is a character sensing device which can feed the foreign language directly into the computer. CIA interest in mechanical translation, character sensing research, and other applications of mechanical or electronic equipment to information processing is coordinated with other intelligence services through another inter-agency committee of the United States Intelligence Board. And again, this group is tied in with the non-intelligence community by including a representative from the National Bureau of Standards who is cognizant of work being done in electronic data processing research by and for many Government agencies, private research organizations, and commercial firms.

Currently we are using microphotography and our Intellofax System as previously described. We are also experimenting with MINICARD equipment manufactured by Eastman Kodak and initially developed for the Air Force. This system in essence substitutes a 16 x 30 mm film strip for the IBM index and document storage aperture cards of our present system. MINICARD document indexes are read electronically rather than mechanically as is the case with IBM cards. Other components of the intelligence community have sponsored important developmental work in automatic abstracting and automatic dissemination.

Trends in the field of information processing would appear to support the following conclusions: 1) Channels for procuring publications and techniques for storing the physical document are extensive and well developed. The outlook is for expansion and intensification of present methods; 2) The type of information service coming into being will demand action primarily in preparing reference personnel to give assistance of higher quality than is given today; 3) Specialized schemes will be developed to fit the needs of specialized users where solutions to problems call for a search of the literature; 4) The present and future demands for reference services will lead to increased use of machines where these can be introduced without jeopardizing the performance of essential intellectual operations.